



Research article

Wavelet-based Benjamini-Hochberg procedures for multiple testing under dependence

Debashis Ghosh*

Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO 80045, USA

* **Correspondence:** Email: debashis.ghosh@ucdenver.edu.

Abstract: Multiple comparisons methodology has experienced a resurgence of interest due to the increase in high-dimensional datasets generated from various biological, medical and scientific fields. An outstanding problem in this area is how to perform testing in the presence of dependence between the p-values. We propose a novel approach to this problem based on a spacings-based representation of the Benjamini-Hochberg procedure. The representation leads to a new application of the wavelet transform to effectively decorrelate p-values. Theoretical justification for the procedure is shown. The power gains of the proposed methodology relative to existing procedures is demonstrated using both simulated and real datasets.

Keywords: correlated statistics; Haar wavelet; high-dimensional data; order statistics

1. Introduction

Consideration of high-dimensional data has come to the forefront of current statistical practice because of scientific advancements in genomics, astronomy and imaging. This has led to a reemergence of the study of multiple comparisons procedures. Many authors have recently advocated for control of the false discovery rate (FDR) [1] relative to the traditional familywise type I error (FWER). The Benjamini-Hochberg (B-H) procedure [1] has received much recent attention in that it controls FDR, which is a more liberal error criterion relative to FWER. This can lead to greater power relative to a multiple testing adjustment using the Bonferroni correction.

One issue that has been of extreme interest has been on adapting the Benjamini-Hochberg procedure to accommodate dependence. The original paper of Benjamini and Hochberg [1] assumed that the p-values being tested were statistically independent. In followup work, Benjamini and Yekutieli [2] showed that the B-H procedure was valid in finite-samples under a condition known as positive regression dependence. More generally, recent work on FDR-controlling procedures under

Table 1. Outcomes of n tests of hypotheses.

	Fail to reject null	Reject null	
True Null	U	V	n_0
True Alternative	T	S	n_1
	W	Q	n

dependence has fallen into three tracks. The first has been to develop procedures that are valid under the positive regression dependence condition and related conditions [3–5]. The second has been to develop FDR-controlling procedures that will be robust to dependence that vanishes asymptotically [6, 7]. The third, which is more recent, has been to model the dependence and incorporate it directly into the FDR-controlling procedure. This has been done by several authors [8–13]; we will review their proposals in Section 2.2.

In this article, we will develop B-H-type procedures for arbitrary dependence. Our starting point is quite different from the methods listed in the previous paragraph. We will make use of the fact that the B-H procedure sorts p-values; this will lead to a useful characterization of the B-H procedure using spacings [14, 15]. The other innovation that will be employed here are wavelets [16, 17]. While FDR-controlling procedures have been developed in conjunction with wavelet estimators [18, 19], these methods bear absolutely no resemblance to the procedures being proposed here. The structure of this paper is as follows. In Section 2, we give some background on multiple testing, describe the procedure of Benjamini and Hochberg [1] and review previous proposals to accommodating dependence. We also provide a brief overview on wavelets. In Section 3, we recast the B-H procedure using spacing results and motivate the wavelet-based methodology alluded to in the title of the paper. In addition, we demonstrate FDR control asymptotically and discuss adaptive estimation. Some simulation studies illustrating the proposed methodology, along with application to several microarray datasets are given in Section 4. We conclude with some discussion in Section 5.

2. Background and preliminaries

2.1. Multiple testing: setup and definitions

We deal with the problem of testing n null hypotheses H_0^1, \dots, H_0^n . Note that the superscript indicates which hypothesis is being tested, while the subscript refers to the null hypothesis. For these null hypotheses, we have corresponding p-values p_1, \dots, p_n . Suppose that out of the n hypotheses, n_0 are true. We can then use the following 2×2 table to define error metrics for multiple comparisons:

The rows of the table allude to truth (i.e., whether the null or alternative is correct), while the column is based on the outcome of the testing procedure. Based on Table 1, the FWER is defined as $P(V \geq 1)$. Benjamini and Hochberg [1] define the FDR to be

$$FDR \equiv P(Q > 0)E \left[\frac{V}{Q} \mid Q > 0 \right].$$

In addition, they propose the following algorithm for multiple testing with p_1, \dots, p_n :

Box 1. Benjamini and Hochberg (1995) [1] procedure

- (a) Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ denote the ordered, observed p-values.
- (b) Define α to be the target FDR. Find $\hat{k} = \max\{1 \leq i \leq n : p_{(i)} \leq \alpha i/n\}$.
- (c) If \hat{k} exists, then reject null hypotheses $p_{(1)} \leq \dots \leq p_{(\hat{k})}$. Otherwise, reject nothing.

Benjamini and Hochberg [1] demonstrated the remarkable result that the procedure in Box 1 controls the FDR at level α when the p-values are independent and uniformly distributed. In fact, the procedure controls the FDR at level $n_0\alpha/n$. This has sparked interest developing adaptive estimators for n_0 or equivalently, $\pi_0 \equiv n_0/n$ [7, 20, 21].

2.2. FDR control under dependence

In terms of results involving error control of the FDR procedure under dependence, they fall into two classes: those that achieve exact FDR control in finite samples and those that achieve asymptotic FDR control. For the first class of results, Benjamini and Yekutieli [2] showed that the Benjamini-Hochberg procedure has exact error control under positive regression dependence. More recently, Sarkar [3, 4] has developed useful inequalities in order to prove the exact FDR control of generalizations of the Benjamini-Hochberg procedure under positive regression dependence. Blanchard and Roquain [5] and Finners et al. [22] provide proofs of the validity of the B-H procedure and related procedures under various dependence conditions related to those considered by [2].

The second class of results involving FDR procedures are asymptotic in nature. Several authors have identified the equivalence between the B-H procedure with a thresholding rule based on the empirical distribution of the p-values [6, 7, 23]. Based on the equivalence, one can then prove results about asymptotic control of the B-H procedure provided that the empirical cumulative distributions of the true null hypotheses and false null hypotheses converge to the correct quantities. In recent work, Wu [24] developed a weak convergence result about the B-H procedure that held under more general dependence conditions.

More recent work has focused on modelling the correlation between test statistics or p-values in the multiple comparisons problem. Efron [10, 11] developed a methodology in which the correlation between the test statistics is modelled parametrically using a latent factor representation. This has been extended by Fan et al. [25] and Desai and Storey [26]. This approach has the effect of accommodating deviations in the variance of the test statistics for the null hypothesis by adjustment for correlation. Perone Pacifico et al. [8] use Gaussian random field theory results [27] in order to construct envelopes for the false discovery proportion. In this approach, the covariance between test statistics is modelled parametrically through use of a covariance function. Leek and Storey [28] model the data using a so-called ‘dependence kernel’ in order to account for correlation. Schwartzman and Lin [13] formulate a model for the mean and variance for the number of rejections. The latter quantity had a term that incorporated correlation between the p-values. Based on the model, they were able to show that the usual FDR-controlling procedure fails to control the FDR asymptotically. Sun and Cai [9] modelled test statistics as arising from a Hidden Markov model. They derived an optimal classification rule using decision-theoretic ideas that was based on what they term as the local index of significance and has a Bayesian interpretation as a posterior probability of differential expression.

2.3. Wavelet estimation

The proposed methodology in Section 3.2. will require use of wavelet estimation procedures. In this section, we provide a brief review of this methodology. The goal is to model a function $y = g(x) + \epsilon$, where y represents an outcome variable, x a fixed predictor variable, g an unknown function and ϵ an error term. We observe the data (x_i, y_i) , $i = 1, \dots, K$. For the purposes of exposition, we will initially assume that $K = 2^J$ for some integer J .

Wavelets are functions that allow for the decomposition of g by projection onto localized basis functions. We can define an appropriate function sequence as follows. We first define the function ϕ by

$$\phi(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{else.} \end{cases}$$

The function $\phi(x)$ is referred to as the Haar father wavelet. Given ϕ , we define

$$V_0 = \{f : f(x) = \sum_{k \in \mathcal{Z}} c_k \phi(x - k), \sum_{k \in \mathcal{Z}} c_k^2 < \infty\}$$

with \mathcal{Z} denoting the set of nonnegative integers. Now recursively define V_i for $i \geq 1$ by $V_i = \{f : f(x) = g(2x), g \in V_{i-1}\}$. This leads to a sequence of subspaces $\{V_n, n \geq 0\}$ with the following properties. First, the subspaces are nested, i.e. $V_k \subset V_{k+1} \forall k \geq 0$. Second, if we define $V = \bigcup_j V_j$, where the union is taken over all $j \geq 0$, then V is dense in $L_2(\mathcal{R})$. This representation is termed as the multiresolution analysis [16]. By orthogonalizing the sequence $\{V_n, n \geq 0\}$, we arrive at the following representation for any $f \in L_2$:

$$f(x) = \sum_{k=0}^{\infty} \alpha_{0k} \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \beta_{jk} \phi_{jk}(x),$$

where $\alpha_{0k} = \int f(x) \phi_{0k}(x)$ and $\beta_{jk} = \int f(x) \phi_{jk}(x)$, the integrals being taken over the real line. The first set of coefficients are referred to as the scaling coefficients, while the latter are termed the detail coefficients. Intuitively, the former capture the trend that is in the data, while the latter model the residuals or noise after the trend is accounted for. Another property is that $\{\phi_{0k}, \phi_{jk}\}$, $j, k \geq 0$ define an orthonormal basis.

While we have described the multiscale resolution analysis approach assuming the father wavelet is the Haar wavelet, it is obvious that the father wavelet is not smooth. Daubechies [17] argued against the use of such wavelets and proposed an alternative father wavelet that is smooth, compactly supported and almost symmetric. It is referred to in the literature as the symmlet. The symmlet does not have a closed analytical form in contrast to the Haar wavelet. We will use this transform throughout the paper. Computationally, estimation in the wavelet framework proceeds by use of the wavelet transform. Wavelets have several desirable properties. First, they are computationally quite fast. Second, it provides good localization in both the frequency and spatial domains. Finally, by projection onto the orthonormal basis functions, wavelets are effective in decorrelating data. This is the property that is of primary interest to our problem.

An example of wavelet estimators is given in Figure 2, which represents data from financial returns from IBM that can be found in the R package `waveslim`. The first line shows the observed data; the second through fifth lines are the wavelet coefficients for levels one through four.

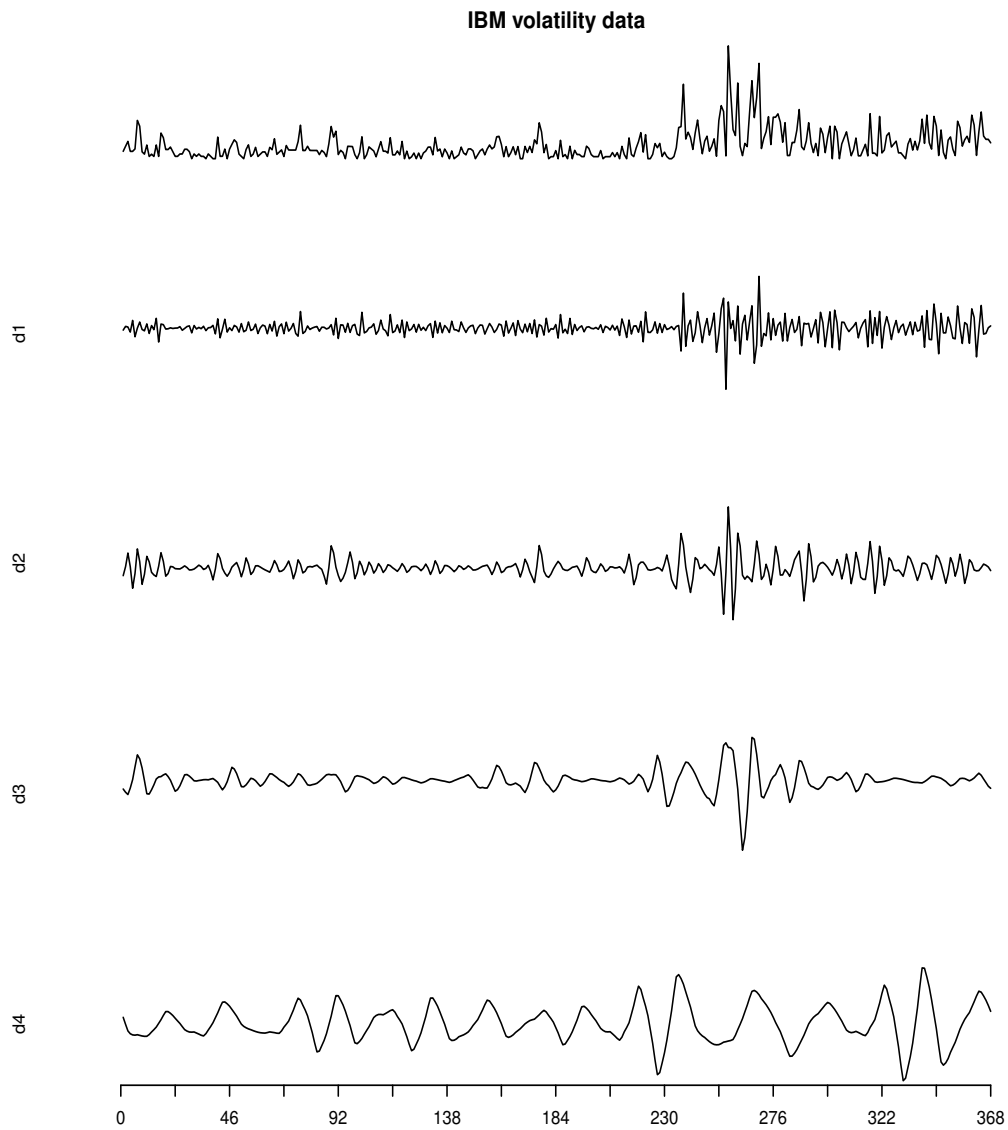


Figure 1. Wavelet decomposition of IBM volatility time series data from the waveslim package in R.

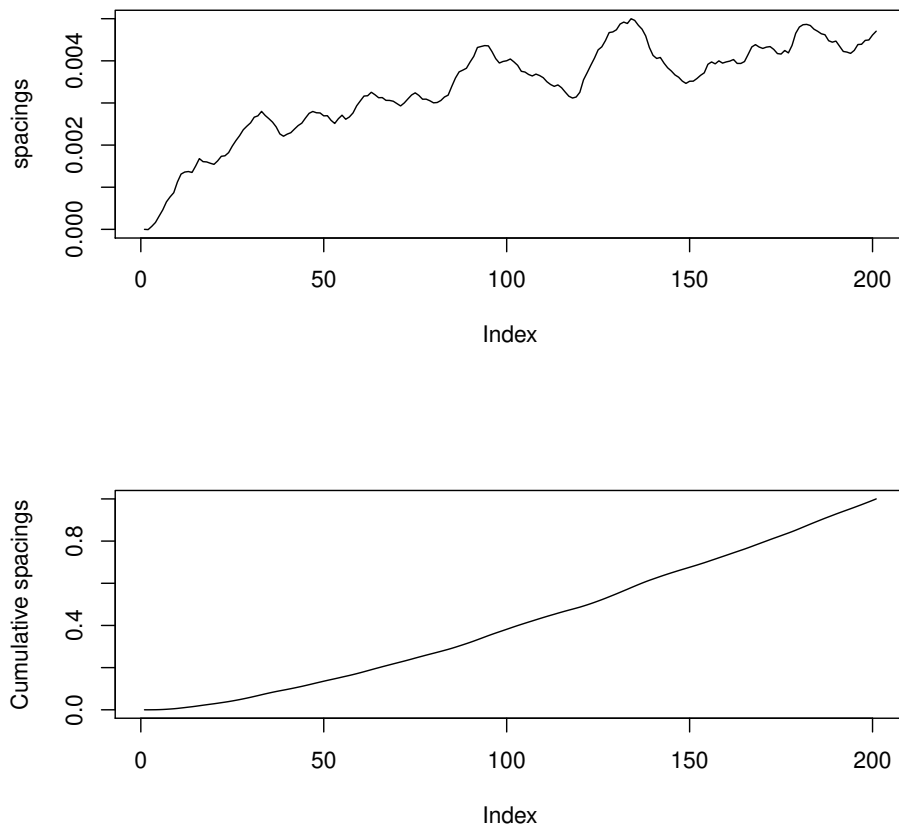


Figure 2. Upper plot shows simulated AR(2) process for the spacings with negative values set to zero. The lower plot shows the cumulative sum process corresponding to upper plot.

As a practical matter, in most situations, the number of tests will not be a dyadic power. In this instance, we will instead use the maximal overlap discrete wavelet transform with a periodic boundary condition [29], which allows for any value of n . The disadvantage is that this is not an orthogonal transformation of the data. In discussing theoretical results in Section 3.3., we will again return to assuming that n is a dyadic power.

3. Proposed methodology

3.1. B-H procedure and spacings

We now review some results from [14] on spacings and show how they can be used to recast the B-H procedure. We will begin by assuming that U_1, \dots, U_n denote a random sample from the Uniform(0,1) distribution. Then the spacings are defined as $V_i = U_{(i)} - U_{(i-1)}$, for $i = 1, \dots, n+1$, where $U_{(i)}$ denotes the i th order statistic of U , $U_{(0)} = 0$ and $U_{(n+1)} = 1$. Note that the joint density of the spacings is given by

$$f(v_1, \dots, v_n) = \begin{cases} n! & \text{if } v_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{n+1} v_i = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

Because of the constraint that the v_i 's sum to one, the spacings are not statistically independent. However, they do have a marginal Beta(1, $n+1$) distribution.

Using these facts, we can reexpress the Benjamini-Hochberg procedure in the following way. Letting

$$\tilde{p}_i = p_{(i)} - p_{(i-1)}, \quad i = 1, \dots, n+1,$$

the B-H procedure in Box 1 rejects $p_{(1)}, \dots, p_{(\hat{k})}$, where

$$\hat{k} = \max\{i : i^{-1} \sum_{j=1}^i \tilde{p}_j \leq \alpha(n+1)n^{-1}E(\tilde{p}_1)\}, \quad (3.2)$$

with $\hat{k} = 0$ if the set in (3.2) is empty.

The B-H procedure assesses clustering of p-values using first-order differences in the sorted p-values. In particular, gaps between the sorted p-values that are smaller than expected, i.e. $E(\tilde{p}_1)$, constitute evidence against the null hypothesis.

There are several insights that are gained by the spacings approach. First, while the p-values for H_0^1, \dots, H_0^n have a certain joint distribution, it is in fact the joint distribution of the order statistics, or equivalently, the spacings that matters in determining the operating characteristics of the B-H procedure. The second insight is that the order statistics, by definition, provides an ordering of the p-values. However, it comes with the constraint that they are in increasing order. By contrast, there is no such constraint for the spacings, but they still have an ordering. The ordering is isomorphic to a one-dimensional time series. It is precisely this property we will exploit when developing our proposed methodology for multiple testing in the presence of dependence.

3.2. New multiple testing procedure

The proposed methodology is conceptually simple in nature and can be described as follows.

Box 2. Proposed procedure

- (a) Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ denote the ordered, observed p-values.
- (b) Define α to be the target FDR. Calculate $\tilde{p}_1, \dots, \tilde{p}_{n+1}$.
- (c) Transform $\tilde{p}_1, \dots, \tilde{p}_{n+1}$ into observations $\tilde{p}_1^*, \dots, \tilde{p}_{n+1}^*$ using the wavelet transform.
- (d) Find $k^* = \max\{i : i^{-1} \sum_{j=1}^i \tilde{p}_j^* \leq \alpha/n\}$.
- (e) If k^* exists, then reject null hypotheses corresponding to $p_{(1)} \leq \dots \leq p_{(k^*)}$. Otherwise, reject nothing.

The key step of the procedure in Box 2 is the transformation in step (c). It has to take the spacings and transform them into new data points that we term *pseudospacings*. To compute pseudo-spacings, we will utilize the wavelet methods discussed in Section 3.1. The code, available at <https://github.com/ghoshd/WaveBH>, shows that we use a non-decimated discrete wavelet transform using the spacings where we only consider the level-1 coefficients. The pyramid algorithm [16] was used in the computation of the discrete wavelet transform. In our simulation studies in Section 4.1., we found these procedures to have satisfactory properties compared to existing methods. Optimization of these parameters from a practical and theoretical perspective is an open question and one that we leave to future research. The strategy we are taking here can be thought of as transforming the problem into a domain that is better behaved statistically, performing the problem there and backtransforming. What we mean by “better statistical behavior” is that by application of the wavelet transform to the spacings, we will be dealing “nearly statistically independent” observations due to their being projections onto orthonormal basis functions. Applying the cumulative sum operator to the spacings will lead to a B-H procedure on transformed order statistics of “nearly independent” p-values. For these p-values, we can simply apply the usual B-H procedure. We will attempt to formalize this notion more mathematically in the next section.

Remark 1. If we are to compare step (d) of the algorithm in Box 2. with (3.2), we see that the difference is we have replaced the spacings for the observed p-values with orthogonal transformations thereof. Note that the orthogonal transformation should have no effect on the mean structure of the procedure, so we would expect $E(\tilde{p}_1)$ and $E(\tilde{p}_1^*)$ to be the same. However, the orthogonal transforms have the effect of transforming the spacings into a space where the correlation structure is lower than for the observed spacings.

Remark 2. An obvious question might be why we are not attempting to apply the wavelet transform to the order statistics of the p-values themselves. Note that if the the original p-values are iid Uniform(0,1) random variables, then the marginal distribution for the i th order statistic will be $Beta(i, n - i + 1)$, $i = 1, \dots, n$. In particular, the i th order statistic has a marginal variance that depends on i . We will use a working regression model to study the behavior of the proposed procedure from a theoretical point of view, and one of the key assumptions will be constant variance of the error term. Such an assumption would be violated by consideration on the original order statistic scale.

Remark 3. In order to develop some intuition behind this procedure, we recall the work of [7]. In their Lemma 3, they show that in the situation of the null p-values being independent, the process

$$\frac{\sum_{i \in \mathcal{N}} I(p_i \leq t)}{t} \equiv \frac{\sum_{i=1}^n I(p_{(i)} \leq t, i \in \mathcal{N})}{t}$$

where \mathcal{N} represents the indices for the true null hypotheses, is a martingale with respect to the σ -filtration $\mathcal{F}_t = \sigma(I\{p_i \leq s\}, t \leq s \leq 1, i = 1, \dots, n)$. The application of the wavelet transform in our

step (c) can be viewed as an attempt to compute orthogonal martingale increments for the process here. Again, the key property that we are exploit of wavelets is the fact that they represent orthogonal projections.

Remark 4. As pointed by a reviewer, the choice of orthogonal basis is relatively arbitrary. We have also implemented the methodology using a Fourier basis, but we found the methodology to have substantially lower power in the simulation studies described in Section 4. How to select an optimal basis onto which to project the spacings remains an open topic for future investigation.

3.3. Theoretical results

As alluded to earlier, we will now assume that n , the number of tests, is a dyadic power in this section. To provide theoretical justification for the proposed procedure, we first consider the case of independent p-values. As noted previously, the spacings will be distributed as dependent Beta(1, $n + 1$) random variables. Then it can be shown through some calculation that the covariance between any two spacings \tilde{p}_i and \tilde{p}_j , $i \neq j$ is given by

$$\text{Cov}(\tilde{p}_i, \tilde{p}_j) = -\frac{1}{(n+1)^2(n+2)}$$

and that the variance of \tilde{p}_i is $n(n+1)^{-2}(n+2)^{-1}$. Thus, the spacings corresponding to a random sample of p-values can be conceptualized as a time series with exchangeable correlation structure between any two time points. While this is certainly a nonstandard type of time series, we will consider other dependence structures later in the section.

In the case of independent p-values, we have that the B-H procedure controls the FDR at level α . It can be reexpressed in terms of the following lemma based on our spacings representation [15]:

Lemma 1. Suppose that the p-values p_1, \dots, p_n correspond to a random sample from the Uniform(0, 1) distribution and that we reject hypotheses $p_{(1)}, \dots, p_{(\hat{k})}$ where

$$\hat{k} = \max\{1 \leq i \leq n : i^{-1} \sum_{j=1}^i \tilde{p}_j \leq \alpha/n\}.$$

Note that if the set is empty, we fail to reject any hypotheses. Then the FDR of the procedure is controlled at level α .

If we apply the wavelet transformation \mathbf{W} to the spacings, then this will not change the joint distribution of the spacings because the wavelet estimator is an orthogonal transformation. Thus, in theory we could replace $\tilde{p}_{(i)}$ by $\tilde{p}_{(i)}^*$, $i = 1, \dots, n$ in the statement of Lemma 1.

We now invoke a working regression model that captures the essential features of this problem. The model is given by the following:

$$\tilde{p}_j = f(j/n) + \epsilon_j \tag{3.3}$$

where f is an unknown function, and ϵ is an error term normally distributed with zero mean and constant variance. Matching this up with the case of the p-values coming from a random sample, we have that $f(x) = (n+1)^{-1}$ and that the variance of ϵ_j is $n\{(n+1)^2(n+2)\}^{-1}$. While a Gaussian error structure cannot be strictly true for the spacings, it nevertheless is useful in finding insights about the proposed multiple testing procedure.

We first consider the case of (3.3) with ϵ_j , $j = 1, \dots, n$ coming from a stationary Gaussian process with covariance $r_l = \text{Cov}(e_j, e_{j+l})$. We next prove the following result.

Theorem 1. Assume that

$$\sum_{l=-\infty}^{\infty} |r_l| < \infty.$$

Then the proposed procedure in Box 1 asymptotically controls the FDR at level α .

Proof. Let $[a]$ denote the greatest integer less than or equal to a . We first define the process

$$\begin{aligned} \tilde{P}_n(t) &= \frac{1}{n} \sum_{j=1}^{[nt]} \tilde{p}_j = n^{-1} \sum_{j=1}^{[nt]} f(j/n) + n^{-1} \sum_{j=1}^{[nt]} e_j \\ &= F_n(t) + \epsilon_n(t). \end{aligned}$$

Then by Lemma 5.1. of Taqqu [33], $n^{1/2}\{\tilde{P}_n(t) - F_n(t)\}$ converges in distribution to $\tau B(t)$, where $B(t)$ is a Brownian motion and $\tau^2 = \sum_{l=-\infty}^{\infty} r_l$. Now define $\eta = \tau n^{-1/2}$. Then $\tilde{P}_n(t)$ can be approximated by

$$\tilde{P}(t) = F(t) + \epsilon B(t), \quad (3.4)$$

where $F(t) = \int_0^t f(s)ds$. Now define $\psi_{jk}(t) \equiv 2^{j/2}\psi(2^j t - k)$ as a wavelet basis on the real line using a wavelet ψ with compact support. Define $\{\tilde{\psi}_l, l \in \mathcal{L}\}$ to be the corresponding wavelet basis on $[0, 1]$; one such construction can be found in Cohen et al. (1992) [34]. The set \mathcal{L} is defined by pairs (j, k) , $j \geq j_0, k = 1, \dots, 2^j$ for the wavelet functions and $(j_0 - 1, k)$ for the scaling functions. Next, we form inner products

$$\begin{aligned} y_l &= \int \tilde{\psi}_l d\tilde{P}, \\ \mu_l &= \int \tilde{\psi}_l f, \end{aligned}$$

and $z_l = \int \tilde{\psi}_l dB$. As $n \rightarrow \infty$, we assume that μ_l approaches zero. This leads to the following equivalent model relative to (3.3):

$$y_l = \epsilon z_l, \quad l \in \mathcal{L}.$$

While the distribution of z_l at a given level is not spherically normal, we can argue as in Johnstone and Silverman ([30], p. 337) that this induces only a slight approximation error so that we can use the following model as being equivalent to (3.3):

$$y_l = \epsilon z_l, \quad l \in \mathcal{L}, \quad z_l \stackrel{i.i.d.}{\sim} N(0, 1). \quad (3.5)$$

For a given level j_0 , the process $\sum_{l \in \mathcal{N}} y_l$ is a martingale, so we can apply arguments as in Lemmas 3 and 4 of Storey et al. [7] to prove that the procedure based on rejecting hypotheses $p_{(1)}, \dots, p_{(\hat{k})}$ where

$$\tilde{k} = \max\{1 \leq i \leq n : i^{-1} \sum_{j=1}^i y_j \leq \alpha/n\}.$$

will control the false discovery rate at level α . Because of the equivalence of (3.5) with (3.3), we have the desired result, which corresponds to step (d) of the proposed algorithm. Based on the application

of the inverse wavelet transform, which is an orthogonal transformation, the FDR is controlled. This completes the proof.

The type of dependence being considered in Theorem 1 can be thought of as short-range dependence. However, because the dependence structure is for the spacings of the p-values, this induces a different dependence structure for the p-values themselves that might not be short-range dependence. As an example, the theorem would cover the situation of the spacings being from an autoregressive process of order 2. A simulated example of the spacings and ordered p-values from such a process is shown in Figure 2. On the p-value scale, there is induced long-range dependence as a result of averaging over the locally correlations from the left-hand graph in Figure 2.

We now discuss how the theoretical results could be extended to the case of a more long-range dependence structure in the spacings. Assume that the spacings satisfy fractional Brownian motion:

$$\text{Cov}(\tilde{p}_j, \tilde{p}_k) = \frac{V_H}{2} (|j|^{2H} + |k|^{2H} - |j - k|^{2H}), \quad (3.6)$$

where $V_H = -\Gamma(2 - 2H) \cos(\pi H) [\pi H(2H - 1)]^{-1}$, and H is the so-called Hurst parameter. Note that there is a one-to-one correspondence between H and the decay parameter α if we assume that the autocorrelation function of the spacings is the of the form $Ak^{-\alpha}$ for $0 < \alpha < 1$. It is given by $H = 1 - \alpha/2$. We can then use the arguments in the proof of Theorem 1 to approximate $\tilde{P}_n(t)$ by

$$\tilde{P}(t) = F(t) + \epsilon^\alpha B_H(t),$$

and $B_H(t)$ is a fractional Brownian motion process, i.e. a zero-mean Gaussian process with covariance function given by the continuous analog of (3.6). Then proceeding as in the proof of Theorem 1, we have that an equivalent model is given by

$$y_l = \epsilon^\alpha \lambda_j z_l, \quad l \in \mathcal{L}, \quad (3.7)$$

where $\lambda_j^2 = 2^{-j(1-\alpha)} \tau^2$, $\tau = 2A\{(1 - \alpha)(1 - 2\alpha)\}^{-1}$ and

$$z_l = \lambda_j^{-1} \int \tilde{\psi}_l dB_H.$$

We point out that from Johnstone and Silverman [30], the z_l are not independent but have bounded dependence in that there exists a constant ρ_0 such that $0 < \rho_0 \leq \text{Var}(z_l | z_m, m \neq l) \leq 1$. One can then develop a B-H style rule in the transformed space and back-transform as in our proposed procedure.

3.4. Adaptive estimation under dependence

For many situations in multiple testing, we expect there to be a large number of true null hypotheses. As alluded to earlier, one finding of Benjamini and Hochberg [1] was that their procedure controlled the FDR at level $n_0\alpha/n$ in the setting where the test statistics for testing n hypotheses are independent. This finding suggests the possibility of estimating n_0 from the data to improve power in the B-H procedure. A large amount of literature has been devoted to developing estimators of n_0 or equivalently, $\pi_0 \equiv n_0/n$. A nice overview of estimators of this quantity can be found in [20].

Given the nature of our testing procedure in Box 2, it seems natural to consider the following adaptive version of the procedure under dependence:

Box 3. Proposed adaptive procedure

- (a) Let $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ denote the ordered, observed p-values.
- (b) Define α to be the target FDR. Calculate $\tilde{p}_1, \dots, \tilde{p}_{n+1}$.
- (c) Transform $\tilde{p}_1, \dots, \tilde{p}_{n+1}$ into observations $\tilde{p}_1^*, \dots, \tilde{p}_{n+1}^*$ using the wavelet transform.
- (d) Estimate n_0 using the algorithm of [21].
- (e) Find $\tilde{k}^* = \max\{i : i^{-1} \sum_{j=1}^i \tilde{p}_j^* \leq \alpha/\hat{n}_0\}$.
- (f) If \tilde{k}^* exists, then reject null hypotheses corresponding to $p_{(1)} \leq \dots \leq p_{(\tilde{k}^*)}$. Otherwise, reject nothing.

We make several observations here. First, while we are using the procedure of Benjamini et al. [21] to estimate n_0 in step (d), any procedure that estimates π_0 could be used in step (d) simply by converting to an estimate of n_0 through $\hat{n}_0 = n\hat{\pi}_0$. Second, the only difference between the procedure in Box 3 versus that in Box 2 is the use of an estimator of n_0 for determining the cutoff in step (d). The procedure in Box 2 implicitly assumes that $n_0 = n$. Examining the proof of Theorem 1, it can be shown in the transformed space, the procedure actually controls the FDR at level $(n_0/n)\alpha$. Thus, slight modifications of the proof of Theorem 1 lead to the following result:

Theorem 2. Assume that

$$\sum_{l=-\infty}^{\infty} |r_l| < \infty.$$

Then the proposed procedure in Box 1 asymptotically controls the FDR at level $n_0\alpha/n$.

4. Numerical examples

4.1. Simulation studies

In this section, via simulation we compare the proposed methodologies in this paper to the Benjamini-Hochberg procedure, the Benjamini-Yekutieli correction [2], which are known to be valid under arbitrary dependence. We wished to compare procedures that only involve the univariate p-values. Two competitor we do not show here are the Bonferroni procedure as well as the correlation-adjusted empirical null hypothesis methodology of Efron (2010) [11], there were two drawbacks against it. First, when π_0 is small, the methodology fails to give a numerically stable estimator for π_0 . Second, in situations where π_0 could be estimated, it turned out that the performance of the method was similar to that from the Bonferroni correction, both of which were worse than the methods presented in the Table (data not shown). We considered several correlation scenarios. The one we report on in this paper is the ‘‘clumpy dependence’’ scenario that was also considered in [7]. This corresponds to the situation in genomics where genes act in networks. The null statistics are distributed as standard normal random variables, while the true alternative have normal distributions with mean two and variance one. For these simulations, 3000 test statistics were correlated in groups of ten with correlation of $\pm\rho$. Formally, for $1 \leq k \leq l \leq 10$,

$$\Sigma_{kl} = \begin{cases} 1, & k = l, \\ \rho, & k < l \leq 5, \\ -\rho, & k \leq 5, l > 5, \end{cases}$$

and the covariance of the test statistics is the Kronecker product of Σ with a 300×300 identity matrix. The proportion of true null hypotheses was set to be $\pi_0 = 0.2, 0.5$ and 0.9 . For each simulation setting, 1000 simulated datasets were generated.

We first show the estimated FDRs at two thresholds, 0.05 and 0.01, of the various procedures in Table 2.

Table 2. Estimated false discovery rates for Benjamini-Hochberg (BH), Benjamini-Yekutieli (BY) and proposed methods (Wavelet and Wavelet-Adapt). The numbers in parentheses are the standard errors. The parameter α denotes the target FDR that we will to control, while π_0 represents the proportion of true null hypotheses.

π_0	ρ_0	α	BH	BY	Wavelet	Wavelet-Adapt
0.2	0.2	0.01	0.02(0.01)	0.002(0.0003)	0.009(0.001)	0.0097(0.001)
0.2	0.4	0.01	0.029(0.01)	0.002(0.0003)	0.009(0.002)	0.0098(0.001)
0.2	0.6	0.01	0.04(0.01)	0.002(0.0003)	0.009(0.001)	0.0099(0.001)
0.2	0.8	0.01	0.031(0.01)	0.002(0.0003)	0.009(0.002)	0.0098(0.002)
0.2	0.2	0.01	0.013(0.01)	0.003(0.019)	0.009(0.001)	0.0091(0.001)
0.2	0.4	0.01	0.019(0.01)	0.003(0.0004)	0.019(0.001)	0.0098(0.002)
0.2	0.6	0.01	0.028(0.01)	0.002(0.003)	0.019(0.004)	0.0099(0.002)
0.2	0.8	0.01	0.035(0.01)	0.003(0.001)	0.019(0.002)	0.0097(0.002)
0.5	0.2	0.05	0.06(0.01)	0.001(0.002)	0.024(0.002)	0.039(0.002)
0.5	0.4	0.05	0.059(0.01)	0.001(0.002)	0.024(0.002)	0.039(0.002)
0.5	0.6	0.05	0.06(0.01)	0.001(0.002)	0.024(0.002)	0.039(0.002)
0.5	0.8	0.05	0.054(0.01)	0.001(0.002)	0.024(0.003)	0.039(0.004)
0.5	0.2	0.01	0.02(0.01)	0.002(0.003)	0.005(0.002)	0.006(0.002)
0.5	0.4	0.01	0.02(0.01)	0.002(0.001)	0.005(0.0002)	0.006(0.001)
0.5	0.6	0.01	0.02(0.01)	0.002(0.001)	0.005(0.002)	0.007(0.001)
0.5	0.8	0.01	0.03(0.01)	0.002(0.002)	0.005(0.002)	0.006(0.003)
0.9	0.2	0.05	0.06(0.01)	0.01(0.003)	0.037(0.005)	0.039(0.005)
0.9	0.4	0.05	0.055(0.01)	0.01(0.003)	0.037(0.006)	0.039(0.006)
0.9	0.6	0.05	0.062(0.01)	0.012(0.004)	0.037(0.007)	0.040(0.007)
0.9	0.8	0.05	0.057(0.01)	0.012(0.006)	0.038(0.009)	0.041(0.009)
0.9	0.2	0.01	0.016(0.01)	0.003(0.002)	0.003(0.002)	0.004(0.002)
0.9	0.4	0.01	0.021(0.01)	0.003(0.003)	0.004(0.002)	0.004(0.002)
0.9	0.6	0.01	0.019(0.01)	0.003(0.003)	0.005(0.003)	0.005(0.003)
0.9	0.8	0.01	0.027(0.01)	0.003(0.004)	0.004(0.0003)	0.004(0.003)

Several interesting findings arise from this table. For BH, we see that FDR control is not achieved with this clumpy dependence scenario, a finding also noted in Storey (2002). As is well-known, the Benjamini-Yekutieli procedures are quite conservative in their control of FDR. In this regard, the proposed methods (the ‘Wavelet’ and ‘Wavelet-Adapt’ columns) tend to be much closer to the nominal FDR levels. They tend to do better for smaller values of π_0 , although there is a suggestion of anticonservatism of the adaptive method for $\pi_0 = 0.5$. The results are all not sensitive to the correlation level, which is an interesting finding. This is in keeping with the idea that the wavelet

transform is effective at removing correlation between the p-values. Next, we study the corresponding powers, shown in Table 3.

Table 3. Estimated powers for BH, BY and proposed methods.

π_0	ρ_0	α	BH	BY	Wavelet	Wavelet-Adapt
0.2	0.2	0.05	0.99(0.02)	0.63(0.005)	0.98(0.01)	0.99(0.02)
0.2	0.4	0.05	0.99(0.02)	0.63(0.005)	0.98(0.01)	0.99(0.03)
0.2	0.6	0.05	0.99(0.01)	0.63(0.005)	0.98(0.01)	0.99(0.03)
0.2	0.8	0.05	0.99(0.01)	0.63(0.005)	0.98(0.01)	0.99(0.02)
0.2	0.2	0.01	0.92 (0.02)	0.31 (0.005)	0.63 (0.05)	0.88 (0.02)
0.2	0.2	0.01	0.93 (0.01)	0.31 (0.004)	0.63 (0.05)	0.89 (0.02)
0.2	0.2	0.01	0.94 (0.01)	0.31 (0.005)	0.63 (0.05)	0.88 (0.02)
0.2	0.2	0.01	0.92 (0.01)	0.31 (0.004)	0.63 (0.05)	0.89 (0.02)
0.5	0.2	0.05	0.91 (0.01)	0.40 (0.004)	0.74 (0.005)	0.86 (0.006)
0.5	0.4	0.05	0.92 (0.01)	0.40 (0.003)	0.75 (0.005)	0.85 (0.004)
0.5	0.6	0.05	0.91 (0.02)	0.40 (0.004)	0.74 (0.005)	0.86 (0.005)
0.5	0.8	0.05	0.92 (0.02)	0.39 (0.003)	0.75 (0.005)	0.85 (0.004)
0.5	0.2	0.01	0.09(0.002)	0.30(0.003)	0.60(0.005)	0.70(0.006)
0.5	0.4	0.01	0.09(0.002)	0.30(0.003)	0.60(0.006)	0.69(0.007)
0.5	0.6	0.01	0.09 (0.002)	0.30 (0.003)	0.60 (0.006)	0.70 (0.007)
0.5	0.8	0.01	0.09 (0.002)	0.30 (0.002)	0.59 (0.005)	0.69 (0.006)
0.9	0.2	0.05	0.11 (0.001)	0.03 (0.001)	0.05 (0.002)	0.05 (0.002)
0.9	0.4	0.05	0.12 (0.001)	0.03 (0.001)	0.05 (0.002)	0.05 (0.002)
0.9	0.6	0.05	0.12 (0.001)	0.03 (0.001)	0.05 (0.002)	0.05 (0.002)
0.9	0.8	0.05	0.12 (0.001)	0.03 (0.001)	0.05 (0.001)	0.05 (0.002)
0.9	0.2	0.01	0.016 (0.001)	0.015 (0.001)	0.012 (0.002)	0.012 (0.002)
0.9	0.4	0.01	0.018 (0.001)	0.014(0.001)	0.012(0.002)	0.012 (0.002)
0.9	0.6	0.01	0.018 (0.001)	0.015(0.001)	0.012 (0.002)	0.012 (0.002)
0.9	0.8	0.01	0.019 (0.001)	0.015(0.001)	0.012 (0.002)	0.012 (0.002)

In terms of power, the proposed methods appear to have slightly larger power than B-Y. While the B-H has much power, keep in mind that it violated FDR control in Table 2. Thus, the comparison with our proposed methodology is not a fair one. In this simulation setup, the main determinant of the power gain appears to be α and π_0 . For smaller values of π_0 , the proposed methods have bigger power gains relative to existing methods. This advantage diminishes as $\pi_0 \rightarrow 1$. Again, the methodology is quite resistant to the correlation structure.

4.2. Gene expression datasets

We applied the proposed methodology to several gene expression datasets. The datasets came from the LBE library in R. The first was the breast cancer gene expression dataset from Hedenfalk et al. [31]. While there were three types of breast cancer samples that were profiled, we focus here on samples with the BRCA1 mutation (seven samples) and those with the BRCA2 mutation (eight samples). After microarray preprocessing, there were 3170 genes under consideration. At a false discovery rate of

0.05, the Bonferroni procedure selected two genes as significant. The Benjamini-Yekutieli (2001) procedure selected three genes as significant [2]. The Benjamini-Hochberg procedure selected 162 genes as significant. The proposed methods in Boxes 2 and 3 selected 92 and 93 genes, respectively, as statistically significant at a false discovery rate level of 0.05.

The next dataset that was studied was the acute leukemia (AML/ALL) dataset from the Golub et al. [32] gene expression study. In this study, there were $n = 7129$ p-values considered, calculated from unpooled two-sample t-tests comparing the samples in the AML group (11 samples) to those in the ALL group (27 samples). Using an FDR of 0.05, the Bonferroni adjustment selected 243 genes, the Benjamini-Yekutieli (2001) procedure 527 [2], Benjamini-Hochberg (1995) 1519 [1], and the procedures in Boxes 2 and 3 1224 and 1331 genes, respectively.

5. Discussion

In this paper, we have developed a B-H-type procedure that accommodates dependence through the use of the wavelet transform. This is done via a spacings decomposition of the B-H procedure [15] to which the wavelet transform is applied. The proposed methodology attempts to keep the structure of the B-H procedure while at the same time decorrelating dependence using the wavelet transform. The intuition behind the procedure is that it attempts to compute orthogonal increments in the wavelet space so that the procedure in the transformed space works with ‘effectively independent’ data. We have shown the superiority of the proposed methodology in the ‘clumpy dependence’ model that was studied in [7]. This situation has proven to be challenging to most multiple testing procedures.

There are many potential extensions of this work. First, the wavelet procedure was not optimized in any sense. We simply used the information from the level-1 wavelet transform. It might be possible to combine information from multiple levels in order to improve power of the procedure. Second, the wavelet basis function might not make sense because of the restriction of spacings to the $[0,1]$ range. While we did some comparisons with using Fourier basis functions, another open problem is to see whether or not use of a different orthonormal basis might lead to different performance of the procedure. The procedure is quite scalable to handling large numbers of p-values. Finally, the intuition behind the procedure that we transform the problem to one where there is ‘independent information.’ This elementary principle should be transferrable to other problems in statistics. We have made available the code for implementing functions and their application to the real datasets in Section 4.2. in github (<https://github.com/ghoshd/WaveBH>).

Conflict of interest

The author declares there is no conflict of interest.

References

1. Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *R. Stat. Soc. B*, **57** (1995), 289–300.
2. Y. Benjamini and D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, *Ann. Stat.*, **29** (2001), 1165–1188.

3. S. K. Sarkar, Some results on false discovery rates in stepwise multiple testing procedures, *Ann. Stat.*, **30** (2002), 239–257.
4. S. K. Sarkar, False discovery and false nondiscovery rates in single-step multiple testing procedures, *Ann. Stat.*, **34** (2006), 394–415.
5. G. Blanchard and E. Roquain, Two simple sufficient conditions for FDR control, *Electron. J. Statist.*, **2** (2008), 963–992.
6. C. R. Genovese and L. Wasserman, A stochastic process approach to false discovery control, *Ann. Stat.*, (2004), 1035–1061.
7. J. D. Storey, J. E. Taylor and D. Siegmund, Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, *J. R. Stat. Soc. B*, **66** (2004), 187–205.
8. M. P. Pacifico, C. Genovese, I. Verdinelli, et al., False discovery control for random fields, *J. Am. Stat. Assoc.*, **99** (2004), 1002–1014.
9. W. Sun and T. Cai, Large-scale multiple testing under dependency, *J. R. Stat. Soc. B*, **71** (2009), 393–424.
10. B. Efron, Correlation and large-scale simultaneous significance testing, *J. Am. Stat. Assoc.*, **102** (2007), 93–103.
11. B. Efron, Correlated z-values and the accuracy of large-scale statistical estimates, *J. Am. Stat. Assoc.*, **105** (2010), 1042–1055.
12. J. T. Leek and J. D. Storey, Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis, *PLoS Genet.*, **3** (2007), e161.
13. A. Schwartzman and X. Lin, The effect of correlation on false discovery rate estimation, *Biometrika*, **98** (2011), 199–214.
14. R. Pyke, Spacings (with discussion), *J. R. Stat. Soc. B*, **27** (1965), 395–436.
15. D. Ghosh, Incorporating the empirical null hypothesis into the Benjamini-Hochberg procedure, *Stat. Appl. Genet. Mol. Biol.*, **11** (2012).
16. S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Machine Intell.*, **11** (1989), 674–693.
17. I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.
18. F. Abramovich and Y. Benjamini, Adaptive thresholding of wavelet coefficients, *Comput. Stat. Data Anal.*, **22** (1996), 351–361.
19. X. Shen, H. C. Huang and N. Cressie, Nonparametric hypothesis testing for a spatial signal, *J. Am. Stat. Assoc.*, **97** (2002), 1122–1140.
20. M. Langaas, B. H. Lindqvist and E. Ferkingstad, Estimating the portion of true null hypotheses, with application to DNA microarray data, *J. R. Stat. Soc. B*, **67** (2005), 555–572.
21. Y. Benjamini, A. M. Krieger and D. Yekutieli, Adaptive linear step-up procedures that control the false discovery rate, *Biometrika*, **93** (2006), 491–507.
22. H. Finner, T. Dickhaus and M. Roters, On the false discovery rate and an asymptotically optimal rejection curve, *Ann. Stat.*, **37** (2008), 596–618.

23. J. A. Ferreira and A. H. Zwinderman, On the Benjamini- Hochberg method, *Ann. Stat.*, **34** (2006), 1827–1849.
24. W. B. Wu, On false discovery control under dependence, *Ann. Stat.*, **36** (2008), 364–380.
25. J. Fan, X. Han and W. Gu, Estimating false discovery proportion under arbitrary covariance dependence (with discussion), *J. Am. Stat. Assoc.*, **107** (2012), 1019–1048.
26. K. H. Desai and J. D. Storey, Cross-Dimensional Inference of Dependent High-Dimensional Data, *J. Am. Stat. Assoc.*, **107** (2012), 135–151.
27. R. J. Adler and J. E. Taylor, *Random Fields and Geometry*, New York: Springer, 2007.
28. J. T. Leek and J. D. Storey , Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis, *PLoS Genet.*, **3** (2007), e161.
29. D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*, Cambridge: Cambridge University Press, 2000.
30. I. M. Johnstone and B. W. Silverman, Wavelet threshold estimators for data with correlated noise, *J. R. Stat. Soc. B*, **59** (1997), 319–351.
31. I. Hedenfalk, D. Duggan, Y. Chen, et al., Gene-expression profiles in hereditary breast cancer, *New England J. Med.*, **344** (2001), 539–548.
32. T. R. Golub, D. K. Slonim, P. Tamayo, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, **286** (1999), 531–537.
33. M. S. Taqqu, Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **31** (1975), 287–302.
34. A. Cohen, I. Daubechies and J. C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Commun. Pur. Appl. Math.*, **45** (1992), 485–560.



AIMS Press

©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)